THE PREDICTION OF ACADEMIC GRADES WITH A PROJECTIVE TEST OF ACHIEVEMENT MOTIVATION: I. INITIAL VALIDATION STUDIES

HENRY N. RICCIUTI

Technical Report No. 1

Office of Naval Research Contract Nonr-694(00)
Project Designation NR 151-113

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

June 1954

THE PREDICTION OF ACADEMIC GRADES WITH A PROJECTIVE TEST OF ACHIEVEMENT MOTIVATION: I. INITIAL VALIDATION STUDIES

Henry N. Ricciuti

Technical Report No. 1

Office of Naval Research Contract Nonr-694(00)

Project Designation NR 151-113

Educational Testing Service

Princeton, New Jersey

June 1954

THE PREDICTION OF ACADEMIC GRADES WITH A PROJECTIVE TEST OF ACHIEVEMENT MOTIVATION: I. INITIAL VALIDATION STUDIES

Abstract

This study was concerned with an evaluation of the validity of a projective "picture-interpretation" test of achievement motivation for predicting academic grades. Results based upon groups of high school juniors, college freshmen, and Naval officer candidates, revealed promising relationships only at the high school level between the test and academic grades, corrected for ability differences. The number of words written in the picture-interpretation stories proved to be very nearly as valid a predictor of high school grades as was the achievement motivation score itself.

The validities of all scoring "categories" included in the scoring system were analyzed separately for each picture included in the study. The results permitted the identification of a small number of scoring categories which tended to yield generally promising positive validities, and which might serve as the basis for an abbreviated achievement motivation score. At the same time, however, considerable fluctuation was revealed in the validities of given categories from picture to picture. This led to the decision to select for cross-validation in a subsequent study a smaller number of the most promising pictures, each to be scored only for the specific categories found to be most valid on that particular picture, in either the positive or negative direction. The results also suggested that absence of reasonably suggestive achievement-related content in the stimulus pictures tended to be associated with poor validity.

It is concluded that the achievement motivation test shows encouraging although not immediately practicable validities in predicting high school grades after ability differences are ruled out. Further research on the test as a predictor of school success certainly seems indicated at both the high school and college level. Validation against broader criteria of personal and scholastic achievement beyond course grades as such would appear to be particularly desirable. Additional research is also needed on the identification of the most generally valid scoring categories, the relationship between validity and picture content, and the value of word output as a supplementary measure of achievement drive.

THE PREDICTION OF ACADEMIC GRADES WITH A PROJECTIVE TEST OF ACHIEVEMENT MOTIVATION: I. INITIAL VALIDATION STUDIES[1]

Henry N. Ricciuti

In a recently published book (4), McClelland, Atkinson, Clark, and Lowell present the results of an imposing body of research, conducted over a period of about five years, on the "achievement motive," or "achievement motivation." In the initial phases of their research, these workers developed what appeared to be a promising technique for measuring the strength of an individual's achievement motivation, or drive to achieve at a high level of excellence. This technique involves an analysis of the content of imaginative stories written as interpretations of stimulus pictures, for the presence of various scoring elements or "categories."

Because the measures thus obtained were judged to hold some promise for predicting various practical criteria, a separate research project was initiated by McClelland, Clark, and Morgan at Wesleyan University for the purpose of exploring the "applied" value of the "picture interpretations" test of achievement motivation, primarily in regard to the prediction of academic achievement at the high school and college level. Some of the preliminary findings obtained with high school students have already been reported by Morgan (5). The present paper deals with the results of a more detailed analysis of the material obtained at the high school level, and with the results of validation studies conducted with college freshmen and Naval officer candidates. The analyses were carried out at the Educational Testing Service, using experimental data which had been collected by the Wesleyan group.

More specifically, the present investigation was concerned with the following major objectives: (a) a further analysis of the validity of the achievement motivation total test score in predicting high school grades, applying a slightly revised scoring procedure to some of the material used in Morgan's analyses (5); (b) a determination of the validity of the total test score in the prediction of academic grades for the freshman year at a liberal arts college, and for the course of training at a Naval Officer Candidate School; and (c) an analysis of the validities of the various scoring elements on each of the separate pictures used in the research. The last mentioned analysis was undertaken for the purpose of identifying the most promising scoring categories and pictures, which might then be incorporated into a more refined version of the test.

A fourth question (d) investigated in this study was the role of word output, or number of words written by the student in his imaginative stories. In any content analysis method involving frequency of occurrence of various elements in written productions, a common problem which arises is the possibility that the measures obtained will reflect the length of the written material. Although McClelland, et al. (4) found little or no relationship between the achievement motivation test score and word output, it was felt that the question deserved further inquiry.

Procedure

Subjects

The study was based upon three major groups of subjects, all males, who took various forms of the picture-interpretation test of achievement motivation. These were: (a) 147 juniors in a large academic high school in a New England industrial city; (b) 406 individuals in a Naval Officer Candidate School (OCS); and (c) 180 freshmen in an Eastern private liberal arts college for men.

Test of Achievement Motivation

Description. The test consists of a series of pictures, usually presented as slides, for each of which the subject is asked to write an imaginative story with a plot and characters. The slides are exposed for ten seconds each, with two and one-half minutes being allowed for the story writing. (In the form administered to the college freshmen group, five minutes were allowed for writing.) A more detailed description of the test and the instructions contained on the cover of the test booklet is given by Morgan (5).

The initial theoretical and experimental studies of McClelland, et al. (4) had been carried out with a small number of pictures. However, for the purpose of exploring the value of the test in predicting academic grades, 41 different pictures were selected for study from a larger pool, some of which had been pre-tested on other high school samples. Magazine illustrations constituted the source of most of the pictures. These pictures were combined into five test forms, which were administered to various subject groups as follows[2]:

A. High School Juniors

Group	A1	Forms E (N=58), F (N=50)
	A2	Forms F (N=43), E (N=39)
	A3	Forms G (N=53), E (N=50)

B. Naval Officer Candidates

OCS Group	B1	Form E (N=101)
	B2	Form F (N=99)
	B3	Form G (N=103)
	B4	Form J (N=103)

C. College Freshmen

Group	C1	Form W (N=180)

Scoring. The basic rationale underlying the method of analyzing stories for achievement motivation is described in detail by McClelland, et al. (4), who also give specific instructions and illustrations for using scoring system C. With the use of a greater variety of pictures, however, this scoring system was modified and expanded somewhat for research purposes (3). The resulting procedure (scoring system D) was used by Morgan in his analysis of the high school data (5), and, with some slight revisions (D-2), this system was used in the scoring of all protocols included in the present study. The high school cases were thus rescored with system D-2, which was found to be very similar to system D, as far as total test score is concerned (r = .86 for Form E, groups 1, 2, and 3)[3].

In scoring a story for achievement motivation, the judge first decides whether it contains any "achievement imagery," i.e., thought content indicating concern with achieving at a high level of excellence. If so, he then determines how many other defined scoring elements are present, with the score for that story being simply the total number of scoring categories or elements found (including achievement imagery itself).

The principal scoring categories included in system D-2 are outlined briefly below:

AI Achievement imagery.

E-1 Explicit evaluation of performance ("he did a good job and was soon promoted"). [Does not contribute to total score.]

N+ Need to achieve ("he wants to be an engineer"; "he hopes to win").

N- Need to avoid failure ("he does not want to fail the critical test").

Ga+ Anticipation of future success ("he knows he will solve the problem").

Ga? Concern about future ("he wonders whether he'll make the grade").

Ga- Anticipation of future failure ("he's worried about failing the course").

G+ Positive affect over success ("he is happy about winning the prize").

G- Negative affect over failure ("he is disgusted with himself for failing").

I Instrumental activity, i.e., activity directed toward an achievement goal. Sub-categorized as follows: I_1: "working carefully"; I_2: "working hard"; I_3: "working for a long time"; I_4: "working fast"; I_u: "unmodified" instrumental activity, i.e., just "working."

I +,?,- Used to indicate whether net outcome of instrumental activity is positive (successful), indeterminate, or negative (failure). [Does not contribute to total score.]

Nup Nurturant press ("the foreman helped him to learn the job").

PA Personal asset ("he was an alert and conscientious student").

B Blocks or obstacles which interfere with or slow down goal directed activity. Sub-categorized as: Bp: personal block ("he failed because of his laziness"), Bw: world block ("the storm kept him from getting through on time"), B?: not identifiable as either Bp or Bw.

AT Achievement thema: Scored if story deals entirely, or almost entirely, with an achievement theme, provided any two of the categories from N to I_4 are present.

In the present research, a number of additional story characteristics not included in the above list were identified for research purposes, but these were not used in obtaining total test scores. Each story was

thus checked for the presence or absence of 29 scoring categories, 19 of which can contribute to total score.

Other Measures Utilized

Word Output. A count was made of the total number of words written by each of the subjects in the three high school groups on Form E of the test, and by approximately one-half of the college freshmen group. For officer candidate group B1, the total number of words written by each student for six even-numbered stories out of the total of twelve was also determined.

Ability Measures. For the high school groups, Otis I.Q.'s (Beta test, Form Cm) obtained during the freshman year were available. In the case of the officer candidates, a weighted composite of the Verbal, Mathematics, and Relative Movements tests (V + 2M + RM) from the Navy Officer Classification Battery was used as an over-all ability measure. For the college freshman group, total score on the American Council on Education Psychological Examination (ACE) was the measure of general ability utilized.

Academic Grades. Average grades for the junior year were used as criterion measures for the high school groups, while freshman year grades were available for the college group. For the OCS groups, a measure of academic achievement was obtained by averaging course grades for the twelve-week period of training.

Adjusted Average Grades. The present study was concerned principally with the question of whether the achievement motivation test held any promise for the prediction of academic success. While it is of interest to know whether a measure of motivation is correlated with academic grades as such, it is of considerably greater interest to know whether a measure of motivation helps to account for that portion of grade achievement

which is not attributable to intellectual ability. Consequently, for each individual an adjusted average grade measure (AAG) was obtained by taking the difference between his actual grade achievement and the grade achievement predicted for him on the basis of his ability test score. This AAG measure can be considered, therefore, as representing the person's grade achievement as compared with that of other individuals having the same ability level, or as reflecting the "non-intellectual" components of academic grades. The AAG criterion scores were obtained particularly for use in the detailed analysis of the validity of specific scoring elements on individual pictures.

Collection of Experimental Data

In all groups subjects were told that the testing was for research purposes only, and that the project was being conducted under the sponsorship of the Office of Naval Research. The high school subjects were tested in the spring of their junior year, the OCS groups during their first month at Officer Candidate School, and the college freshmen during the first week of the freshman year, in connection with a "pilot" freshman testing program.

Analyses and Results

The results of the research can be divided into two major areas: (a) findings of those analyses where principal interest was centered upon the validity of the achievement motivation total test score, as well as upon the influence of the word output variable already mentioned; (b) findings of the analyses which were aimed at identifying the more promising scoring categories and stimulus pictures. The results obtained in these two major areas are discussed separately in the sections which follow.

Analyses at the Total-score Level

High School Subjects. Table 1 presents the intercorrelations obtained among achievement motivation scores, word output, Otis I.Q.'s, and the grade achievement measures. These correlations are based upon data for all three high school groups combined, and they involve only Form E of the achievement motivation test. The three groups were combined in order to obtain more stable over-all results. It will be noted that separate correlations are presented for total scores on the Achievement Imagery and Achievement Thema categories as well as for the over-all total score on the test. These two categories were included separately since it has been suggested that they might serve as a basis for an abbreviated achievement motivation score.

The achievement motivation measures are seen to have moderate positive correlations with junior grades (r's from .23 to .33), and considerably lower correlations with intellectual ability (.15 to .17). Of further interest is the finding that when grades are adjusted for differences in ability, the achievement motivation measures continue to show significant positive correlations (.16 to .29) with the non-intellectual components of grade achievement (AAG). The total test score yields a correlation of .26 with AAG, while Achievement Thema alone yields a correlation of .29, which indicates that an abbreviated subscore may be just as effective for some purposes as the considerably more complicated total score.

Although the aforementioned correlations are too small to be of much practical value for predictive purposes, they are regarded as being quite promising in that they reveal a significant relationship between the measure of the achievement motive and that portion of grade achievement not accounted for by intellectual ability. The separate contribution

Table 1

Intercorrelations among Achievement Motivation Test Scores (Form E), Word Output, Otis IQ, and Grade Achievement Measures.
(High school juniors, males. Groups A1,A2,A3, N = 147. Decimals omitted.)

		Achievement Motivation Test				Otis IQ	Avg. Grade	AAG	Mean	σ
		AI	AT	AI+AT	W					
		2	3	4	5	6	7	8		
Ach. Motiv. Test (E) Total Score (12 pictures, 2½')	1	92	86	98	50	17	31	26	13.46 (16.13 (11.15	7.98 9.03)[a] 9.13)[b]
Achievement Imagery	2		66	94	nc	16	23	16	4.63	2.18
Achievement Thema	3			87	nc	15	33	29	1.43	1.45
AI+AT (2 + 3)	4				nc	17	29	23	6.06	3.32
Word output, total	5					14	25	21	512.86	122.96
Otis IQ	6						50	-05	108.69	9.94
Avg. jr. grades	7							84	80.08	7.84
Adjusted Avg. Grades (AAG)	8								12.97	3.75

Note: Correlations $\geq$ .21 and .16 reach the .01 and .05 levels of confidence, respectively.

[a] Form F, Groups A1,A2, N = 93

[b] Form G, Group A3, N = 53. Total N for Forms F and G not identical with total N for Form E since some subjects did not take two forms of test.

[nc] Correlations not computed.

of the total achievement motivation score to the prediction of grades is also evidenced in the fact that the correlation of .50 between I.Q. and grades is raised to .55 when achievement motivation is combined with I.Q. in a multiple correlation.

When attention is directed to the word output variable, it is noted that mere number of words written by the individual on the achievement motivation test correlates .50 with the test score itself, .14 with Otis I.Q., and .21 with AAG. These are findings of considerable interest, since (a) they reveal a higher relationship between length of story and test score than that found by McClelland, et al. (4) (r^2s from .11 to .25), and (b) it is apparent that for all practical purposes, one could predict high school grades very nearly as well by using number of words written as by using the total achievement motivation test score. The latter point is also revealed in the finding that combining word output with I.Q. raises the correlation of I.Q. and grades from .50 to .53, which is almost the same effect produced by combining the achievement motivation score with I.Q. Although word count and achievement motivation each show a significant positive relationship to the non-ability components of grades, when the two are combined in predicting AAG there is virtually no gain over the correlation shown by achievement motivation alone (r with AAG increases from .26 to .27). It might be mentioned, finally, that the correlation of .26 between achievement motivation total score and AAG drops to .18 when word count is partialled out.

Naval Officer Candidates. It is apparent from the data presented in Table 2, that the achievement motivation scores were not related to final academic average at Officer Candidate School. Also, the achievement motivation test scores were even more independent of general ability

Table 2

Intercorrelations among Achievement Motivation Test Scores, Word Output, Navy Officer Classification Battery Scores, and Academic Grades (Navy officer candidates, groups B1, N = 101; B2, N = 99; B3, N = 103. Decimals omitted.)

		Group	Word Output	OCB	Final Academic Average	Mean	σ
Ach. Motiv. Test	Form						
Total Score (12 pictures $2\frac{1}{2}'$)	E	B1	33	03	08	23.09	10.64
	F	B2	nc	05	02	16.85	7.87
	G	B3	nc	07	03	22.16	11.31
Word Output, Total (6 pictures $2\frac{1}{2}'$)	E	B1		-06	-05	303.66	56.53
Officer Classification Battery: Weighted Composite (V + 2M + RM)		B1			54	23.21	3.95
		B2			55	24.88	4.08
		B3			32	24.59	3.65
Final Academic Average, Officer Candidate School		B1				77.23	4.66
		B2				77.84	5.52
		B3				77.96	4.30

Note: Correlations $\geqq$ .25 and .20 reach the .01 and .05 levels of confidence, respectively.

[nc] Correlations not computed

(r^2s from .03 to .07) than was found to be the case in the high school sample (r^2s from .15 to .17). (The validities of the test for predicting adjusted average grades were not determined, since the correlations with grades as such were so low.)

It is interesting to note that for group B1, word output on the achievement motivation test also failed to show any relationship with academic grades. Thus, neither the test score nor story length revealed any validity in predicting grade achievement for the officer candidate groups, as contrasted with the positive relationships found for both these measures in the high school sample. Word output and test score, however, again showed an appreciable positive relationship with one another ($r = .33$).

College Freshmen. Approximately one half of the total college group was used in the total score analysis. The results obtained with this sample were very similar to those found for the officer candidate groups. As indicated in Table 3, the achievement motivation test score showed no relationship to average grades for the freshman year, and it was also unrelated to intellectual ability.[4] Similarly, word output was not correlated with grades; its relationship to the test score, moreover, was considerably less ($r = .16$) than had been found in the case of the high school and officer candidate groups.

An examination of the various achievement motivation test score means in Tables 1, 2, and 3 reveals considerable variation for different test forms as well as for different subject groups. Unfortunately, the same judges did not participate equally in the scoring for the different test forms and groups, and since Form W[5] was taken only by the college freshman group, no rigorous systematic comparisons of the obtained means can be made. However, there is some reason to believe that the officer

Table 3

Intercorrelations among Achievement Motivation Test Scores (Form W), Word Output, ACE Scores, and Academic Grades

(College freshmen, males. Group C1, N = 88. Decimals omitted.)

		Word Output	ACE Total	Avg. Grade	Mean	σ
		2	3	4		
Ach. Motiv. Test (W)						
Total score (8 pictures, 5′)	1	16	00	-01	22.52	8.19
Word output, Total	2		07	06	742.50	137.21
ACE Total Score	3			45	64.73	9.33
Average grades, freshman year	4				796.68	67.19

Note: Correlations $\geq$ .27 and .21 reach the .01 and .05 levels of confidence, respectively.

candidate groups achieved higher mean scores than the high school students (at least on Forms E and G). This effect may be due in part to the fact that the officer candidate groups wrote somewhat longer stories (51 words per story for group B1, as compared with 43 words per story for groups A1, 2, and 3).

It may also be noted, by examining the standard deviations of the test scores relative to their respective means, that there is somewhat less variability in the officer candidate groups and in the college freshman group than there is in the high school groups. This is also true in regard to the variability of the word output measures.

Analyses of Scoring Categories and Pictures

One of the major objectives of this study was to identify the particular scoring categories and stimulus pictures which showed the most promising validities for the prediction of grade achievement. Of the 41 pictures included in the pool, some could be expected to be more effective than others. Similarly, for research purposes the stories had been examined for a considerably larger number of scoring elements than one would expect to use in any final form of the test. In approaching this problem, it was decided not to work with total scores on particular scoring categories (based upon all pictures in a given form) or with total scores on particular pictures (based upon all categories). Rather, it was decided to examine the validity of each scoring category on each of the 41 pictures included in the study. Such analyses would thus indicate which scoring categories yielded the most promising validities, and on what pictures; at the same time, for every picture the validities of the various categories used in scoring it would be known.

Separate analyses were carried out for high school, officer candidate, and college freshman subjects, with the criterion in every case being grade achievement, corrected for differences in ability (AAG). For each of the 29 scoring categories, 92 sets of data for different pictures and subject groups were analyzed for evidence of validity.[6] Using the point-biserial correlation between presence or absence of the category and AAG as an index of validity, a determination was first made of those categories having a validity equal to or greater than .10, either positive or negative. For these categories only, the actual biserials were then computed. Thus, for each category on every picture and subject group, the direction of the validity, as well as the magnitude of the validity if it happened to be .10 or greater, were known. (If a catergory had a frequency less than 5 for a particular picture and group, no validity index was computed.)

When the results of the above analyses were studied, one of the most noticeable findings was the variation in both the size and direction of the validity of particular categories from picture to picture. Such picture variation was noticeable not only across different groups of subjects, but also within subject groups. Hence, although the majority of category validities were positive, particularly in the high school groups, a fairly high degree of picture specificity was indicated insofar as category effectiveness is concerned. No meaningful trends relating specific category validities to picture content could be discovered. As expected, some of the pictures yielded a good many scoring categories with appreciable positive validities, while others produced few such categories, or yielded both positive and negative category validities.

In view of these findings, it was decided to select for cross-validation in another study those pictures which generally yielded a good

many categories with relatively high validities, whether these validities were positive or negative. In the cross-validation study which was to follow, each picture would then be scored only for those categories found to have a validity equal to or greater than .10, and in the direction indicated by the empirical results of the present investigation. A group of eighteen pictures was therefore selected for this purpose. The details and results of this cross-validation study, which was conducted at the high school level, are described in another paper (7).

Picture Content and Validity. The data revealed some interesting and suggestive preliminary findings as to the relation between over-all validity for predicting grade achievement and gross differences in picture content. It was possible to classify the 41 pictures used in the study into four somewhat coarse groups, on the basis of the apparent immediate relevance of the picture content to academic or "white collar" achievement or task situations, as follows:

A. "White collar" group -- twelve pictures showing a man or youth working at a desk, in school, in a laboratory, etc.

B. "Mechanical-physical" group -- eleven pictures showing a person or persons working at a machine, or in a factory, or engaged in some activity such as skiing, driving a truck, performing a trapeze act, etc.

C. "Borderline" group -- nine pictures in which the individual may or may not be viewed as being engaged in an achievement activity, i.e., the situation may be seen as involving primarily interpersonal or social elements.

D. "Unrelated" group -- nine pictures whose content generally contains few or no obvious achievement-related cues (for example, the "father and son" and "boy at window" picture from the Thematic Apperception Test).

Table 4 presents the percentages of positive and negative category validity indices yielded by each of the four types of pictures, as well as the percentages of category validities which were equal to or greater than .10, both positive and negative. These figures are given separately for the high school, officer candidate, and college freshman groups. In each case the total number of category validity indices serving as the base N for the four reported percentages is also indicated, along with the number of pictures involved.

Examination of the percentages in Table 4 indicates that, in general, the "white collar," "mechanical-physical," and "borderline" pictures yielded both a higher percentage of positive category validities and a greater preponderance of positive over negative validities, than was the case for the "unrelated" pictures. This trend is especially noticeable if one considers the validities equal to or greater than .10, where, for officer candidate and college freshman subjects the "unrelated" pictures actually yielded more negative than positive validities. In addition, for the two groups of subjects just mentioned, the preponderance of positive over negative validities equal to or greater than .10 showed a gradual decrease starting from a maximum for the "white collar pictures," through the "mechanical-physical," "borderline," and "unrelated pictures," in that order.

These results suggest that in general, pictures portraying people engaged in white collar, mechanical, or physical tasks or activities tend to yield more scoring categories with appreciable positive relationships to the grade achievement measures (AAG) than is the case for pictures with content of an unrelated or entirely different sort. This trend seems to be especially noticeable in those two subject groups where the test on the whole showed generally poor validity, namely the officer

Table 4

Percentages of Positive and Negative Category Validities Yielded by Four Types of Pictures in Predicting Adjusted Average Grades

GROUPS		Types of pictures: White-collar % of category validities +	White-collar -	Mechanical-physical % of category validities +	Mechanical-physical -	Borderline % of category validities +	Borderline -	Unrelated % of category validities +	Unrelated -	All Pictures % of category validities +	All Pictures -
High School Juniors	Total %	78	18	75	24	85	15	69	31	77	20
	Total % ≥ .10	44	4	46	7	56	6	25	19	46	6
	Total no. validities and pictures	(116) (11)		(109) (11)		(55) (9)		(16) (5)		(296) (36)	
Officer Candidates	Total %	60	34	55	41	62	32	57	39	58	37
	Total % ≥ .10	25	8	20	9	27	15	13	19	22	11
	Total no. validities and pictures	(227) (16)		(198) (13)		(108) (11)		(70) (8)		(603) (48)	
College Freshmen	Total %	56	38			60	25	49	49	54	41
	Total % ≥ .10	23	9	None		15	5	8	11	15	10
	Total no. validities and pictures	(66) (3)		Included		(20) (1)		(71) (4)		(157) (8)	

Note: Total % of positive and negative validities do not always add up to 100 because of occasional zero validities.

candidate and college groups. As expected on the basis of the total score correlations already reported, the percentages of positive and negative validities for all pictures combined reflect the generally more positive results obtained in the high school groups.

Relative Over-all Validities of Scoring Categories. As mentioned previously, considerable fluctuation was found in the validities of various scoring categories when the results obtained with different pictures and subject groups were examined. However, a summary of the over-all validities yielded by the principal scoring categories should be of interest in evaluating the comparative general effectiveness of the various categories when applied to a fairly large number of different pictures and subjects. Table 5 presents, for each of the principal scoring categories, the percentages of obtained validity indices which were positive and negative, as well as the percentages of validities, both positive and negative, which had a point-biserial value equal to or greater than .10. The total number of category validities analyzed, upon which the percentages are based, is also indicated.

In general, the following scoring categories appeared to show the most promising relationships to the adjusted grade achievement measures, in terms of reasonably frequent occurrence, high percentage of positive validities, and relative preponderance of positive over negative validities: AI (achievement imagery), E-1 (explicit evaluation), Ga+ (anticipation of future success), G+ (positive affect over success), I_3 (working for a long time), I+ (outcome of instrumental activity successful), and AT (achievement thema).

It is interesting to note further that virtually all of the scoring categories yielded a higher percentage of positive than negative validities. This was true even for categories like Ga- (anticipation of future failure), G- (negative affect over failure), I- (outcome of instrumental activity unsuccessful), and B (blocks or obstacles), although the positive

Table 5

Percentages of Positive and Negative Validities Yielded by Various Scoring Categories on All Pictures and Groups Analyzed.

	Total no. validities analyzed[a]	Total percentage +	Total percentage -	Total percentage ≥ .10 +	Total percentage ≥ .10 -
AI (achievement imagery)	92	66	26	35	10
E-1 (explicit evaluation)	79	65	28	33	9
N+ (need to achieve)[b]	59	51	46	22	10
Ga+ (anticipation of success)	59	63	32	17	3
Ga? (concern about future)	33	58	39	9	12[c]
Ga- (anticipation of failure)	34	56	41	21	9
G+ (positive affect re success)	57	63	32	26	7
G- (negative affect re failure)	37	54	43	22	8
I (instrumental acts):					
I_1 ("working carefully")	17	41	59[c]	24	6
I_2 ("working hard")	53	47	49[c]	15	9
I_3 ("working a long time")	40	73	27	30	10
I_4 ("working fast")	2	100	0	0	0[c]
I_u (just "working")	56	50	45	18	16
I+ (outcome successful)	57	70	23	40	5
I? (outcome indeterminate)	69	54	41	13	19[c]
I- (outcome unsuccessful)	10	80	10	10	0
Nup (nurturant press)	16	50	50[c]	19	6
PA (personal asset)	33	67	33	27	6
Bp (personal block)	23	70	30	13	9
B? (block, type uncertain)	15	73	27	20	13
Bw (world block)	24	67	25	21	0
AT (achievement thema)	79	62	33	35	6

[a]Maximum number of analyses possible is 92. If a category occurred fewer than 5 times for a particular picture and group analyzed, no validity was computed.

[b]The scoring category "N-" (need to avoid failure) was dropped from the analyses because of rare occurrence.

[c]Comparisons failing to show preponderance of positive over negative validities.

relationships here were not as marked as those yielded by the categories mentioned in the preceding paragraph. Nevertheless, these positive trends are of considerable theoretical interest, since one might expect that the thought content represented by the four categories just mentioned might be more typical of the under-achieving or failing student, if anything, than of the succeeding or over-achieving student.

Such a view appears to have been taken by MacArthur (1), who recently analyzed TAT stories of college over- and under-achievers for various achievement motivation scoring elements defined in an early study by McClelland, et al. (2). Because of an absence of positive trends between over- and under-achievement and four scoring elements closely resembling the Ga-, I-, Blocks, and combined Instrumental Activity categories used in the present study, MacArthur concluded that these scoring elements rested on a weak theoretical basis. However, the results of the present study indicate that such categories may in fact be more typical of the over-achievers than of the under-achievers, even though the trends in this direction are not as marked as is the case for some other categories. The apparent discrepancy between these findings and MacArthur's results may be due to slight differences in the definition and use of the scoring categories, but more probably they are due to picture differences, MacArthur having used standard TAT pictures with content having little immediate relevance to academic achievement-related activities.

Discussion

Test Validity. The validities obtained with the high school groups indicate that the achievement motivation test bears a significant relationship to grade achievement, even when ability differences are ruled out. The independent contribution made by the test when combined with an intelligence measure for predicting academic grades is clearly too small to be of any immediate practical value, particularly when one considers the relative difficulty involved in obtaining the projective test scores. Nevertheless, the positive relationships yielded by the achievement motivation test are considered to be quite encouraging particularly when one considers the relatively low reliability of the present test and the relatively complex and "impure" nature of the AAG measures.

Although we assume that differences in achievement motivation account for some of the variation in grades after ability differences have been ruled out, it is generally acknowledged that a large number of additional factors play a substantial part in determining where individuals of equal ability will fall on a grade continuum. Hence, the positive relationships of the test with this complex criterion appear to be quite promising at the high school level, and should encourage further research in this area. There does not appear to be a clear-cut explanation for the essentially negative results obtained with the officer candidate and college freshman groups. Possible reasons for these contrasting findings are discussed in a subsequent section of the paper.

Word Output. The fact that word output functioned very nearly as well as the achievement motivation score itself in predicting high school grades suggests that the amount written by the individual may possess real significance as an additional indicator of achievement drive. This notion is given further support by the results obtained in the cross-validation study at the high school level (7), as well as in a study comparing the stories written by "relaxed" and "achievement oriented" subjects (6) in which the latter group, presumably more highly motivated, were found to write significantly longer stories. The lack of relationship between word output and grade achievement in the OCS and college groups may be attributable to some of the same factors discussed later in relation to the achievement motivation score itself.

The correlations between word output and achievement motivation test score found in the high school and officer candidate groups (.50, .33) are considerably higher than those (.11, .25) reported by McClelland, et al. (4). These higher correlations may be due in part to the shorter time limits for each story (two and one-half minutes), which may produce greater differences in the amount written by highly and poorly motivated students. This explanation would be consistent with the low correlation (.16) found in the present study for the college freshman group, where the time limit was five minutes per picture. The theoretical and practical significance of word output, or over-all productivity, on certain types of tasks (such as writing picture interpretation stories) certainly seems to be worth a good deal of additional study.

Picture Content. The results of the analysis at the picture level indicate strongly that picture content is a very important variable in the measurement of achievement motivation through the picture interpretation approach. This conclusion is indicated by the picture-to-picture variation in the validities of given scoring categories, and by the results suggesting a positive relationship between over-all test validity and the presence of achievement-related content in the stimulus pictures. These findings lend some support to the notion already suggested by McClelland, et al. (4, pp. 213-214) in their discussion of "picture cues and motive extensity," that the best predictions of different criteria may be yielded by pictures whose content shows some obvious relevance to the particular type of behavior to be predicted. Certainly this whole problem, which essentially involves the question of the generality of the achievement motivation measures, represents a critical area needing further systematic research.

Scoring Categories. The over-all validities yielded by the various scoring categories indicate that a small number of them seem to yield generally positive validities on a great variety of pictures. This suggests that to obtain a rough measure of achievement motivation using a variety of pictures, a score based primarily on a few of the categories listed might be quite useful. As a matter of fact, the total score correlations for the high school students indicate that using the Achievement Thema category alone produces approximately the same validity as that of the over-all score based on all categories. The question of whether or not one attains highest validity by scoring each picture for those specific categories found to be best on that picture, was investigated in the cross-validation study already mentioned (7).

Discrepancy between Results at High School and College Levels. In view of the positive findings obtained at the high school level, the essentially negative results found in the officer candidate and college freshman groups are difficult to explain. It may be that the college and OCS subjects represent highly selected groups motivationally, having much less "real" variation in achievement drive than is found among high school students. However, the somewhat lower achievement motivation variabilities found in the older groups are not so markedly reduced that they alone could have accounted for the virtually zero validities obtained.

It is possible that there may have been differences in the attitudes with which students in the various groups approached the test, a factor which is known to have noticeable effects on the test results. Although all groups took the test under presumably "neutral" conditions, involving only the achievement cues normally present in a classroom situation, with the test being presented "for research purposes," it may be that under these conditions the test differentiated more validly among individuals at the high school level than at the older levels. Still another factor possibly contributing to the poor validities at the college level might be the inadequacy of the particular stimulus pictures which were used with these groups.

It is conceivable, also, that in the high school situation more than at the college level, attitudinal and motivational characteristics shown by the student in his overt behavior may be reflected in the course grades assigned to him by teachers. If so, this might raise the validities of the achievement motivation test in predicting high school grades.

A further possibility of particular interest is that at the high school level, strength of achievement drive as such, i.e., as a characteristic trait of the individual, has a more direct and straightforward

influence upon the grade a student attains, relative to his ability level. In some college situations, for example, many students with strong achievement motivation may gradually learn to settle for somewhat lower course grades than they would have been satisfied with at the high school level, if there are strong social pressures making it somewhat undesirable to attain extremely high marks, or if the student begins to place particularly high value upon achievement in extra-curricular activities at the expense of course grade achievement. On the assumption that such factors are more influential at the college level than at the high school level, they would tend to lower the correlations between college grades and achievement motivation measures.

On the other hand, in some college situations such as the freshman college year, perhaps, and particularly in an officer training school setting, the situational demands for high grade achievement may be so strong that individuals who characteristically have low achievement motivation may be stimulated to achieve considerably higher grades than they normally would attain in a situation possessing fewer externally generated achievement demands. If these increased situational pressures affecting grade performance are not correspondingly revealed in the individuals' test stories as well, this too would tend to lower the correlations between measures of "internal" achievement motivation and grades.

It is obvious that further research at both the high school and college level is needed before the questions discussed in the preceding paragraphs can begin to be answered. If one is interested primarily in evaluating the test as a measure of achievement motivation, then the critical problem is that of securing the most relevant criteria to serve as standards against which the test can be validated. Although attempts

to improve the prediction of academic grades through the use of tests of personality and motivation are well worth undertaking for practical purposes, it should be kept in mind that for reasons already indicated, adjusted average grade measures are of questionable theoretical value as criteria for judging whether a test measures achievement drive.

Furthermore, it is generally agreed that there are many important aspects of student growth and achievement which are not reflected directly in grade point average. Many educators have been concerned with the definition and measurement of these broader criteria of scholastic and personal attainment in various areas, including extra-curricular activities, for example. It would be highly desirable to validate tests of motivation, such as the one involved in this study, against these broader criteria of achievement which go beyond mere classroom performance. (Exploratory studies along these lines have been initiated by Dr. D. C. McClelland at Wesleyan University.) Perhaps particularly meaningful behavioral criteria of achievement motivation might be secured by examining the extent to which the individual undertakes achievement-directed activities, or the extent to which he actually achieves, in areas where his behavior is influenced more by spontaneous "internal" motivation than by situational, external pressures toward achievement.

Another approach to this problem of validity has been taken in a separate study (6) in which a sample of twelve pictures and all of the present scoring categories have been evaluated in terms of their effectiveness in discriminating between two different groups of subjects, one tested under "relaxed" conditions, the other under "aroused" conditions intended to elicit experimentally a higher degree of achievement motivation. This "theoretical" criterion represented by the contrasting experimental conditions is essentially that used by McClelland, et al. (4) in

their original theoretical work on the development of the achievement motivation test.

Summary

This study was concerned with an evaluation of the validity of a projective "picture-interpretation" test of achievement motivation for predicting academic grades. Results based upon groups of high school juniors, college freshmen, and Naval officer candidates, revealed promising relationships only at the high school level between the test and academic grades, corrected for ability differences. The number of words written in the picture-interpretation stories proved to be very nearly as valid a predictor of high school grades as was the achievement motivation score itself.

The validities of all scoring "categories" included in the scoring system were analyzed separately for each picture included in the study. The results permitted the identification of a small number of scoring categories which tended to yield generally promising positive validities, and which might serve as the basis for an abbreviated achievement motivation score. At the same time, however, considerable fluctuation was revealed in the validities of given categories from picture to picture. This led to the decision to select for cross-validation in a subsequent study a smaller number of the most promising pictures, each to be scored only for the specific categories found to be most valid on that particular picture, in either the positive or negative direction. The results also suggested that absence of reasonably suggestive achievement-related content in the stimulus pictures tended to be associated with poor validity.

It is concluded that the achievement motivation test shows encouraging although not immediately practicable validities in predicting high school

grades after ability differences are ruled out. Further research on the test as a predictor of school success certainly seems indicated at both the high school and college level. Validation against broader criteria of personal and scholastic achievement beyond course grades as such would appear to be particularly desirable. Additional research is also needed on the identification of the most generally valid scoring categories, the relationship between validity and picture content, and the value of word output as a supplementary measure of achievement drive.

Footnotes

1 This research was supported by the Office of Naval Research, initially under Contract Nonr-559(03) with Wesleyan University, and subsequently under Contract Nonr-694(00) with the Educational Testing Service. The selection and pre-testing of the pool of stimulus pictures, collection of experimental data, and a large part of the test scoring were carried out at Wesleyan University principally by Doctors D. C. McClelland, R. Clark, and H. Morgan. Some of the scoring and all of the analyses reported herein were conducted at the Educational Testing Service.

2 Forms E, F, and G each consisted of twelve non-overlapping pictures, while Form J represented a combination of twelve pictures taken from E, F, and G. The remaining five pictures in the pool of forty-one, along with two pictures from Form E and one from Form G, constituted Form W. Each of the high school groups took two different forms of the test with a five week intervening period. (Alternate form reliabilities for these groups reported by Morgan (5) were .56, .56, and .64.) In order to increase the stability of the results at the high school level, analyses were carried out using all data available for a given form, i.e., 147 cases for Form E, 93 cases for Form F, and 53 cases for Form G.

3 Scorers participating in the study included H. Morgan, R. Clark, and R. Lull at Wesleyan University and H. Ricciuti and D. Saunders at Educational Testing Service. Judges did not participate equally in the scoring of all groups. In a separate study (8) involving one of the above scorers (H. Ricciuti) and two other judges, it was found

that high inter-scorer agreement can be attained for System D-2 as indicated by total score correlations among the three judges of .81, .83, and .74. For comparative purposes the 30 illustrative protocols scored by McClelland, et al. (4) with Scoring System C were rescored with System D-2. Correlations between C and D-2 were .84 for total score, and .79 for a sub-score based upon "Achievement Imagery" and "Achievement Thema" alone.

4 An unpublished analysis of the entire sample by Dr. H. Morgan using Scoring System D yielded a correlation of .07 between total achievement motivation score and freshman year grades.

5 Each page in the test booklet for Form W was divided into fourths by four questions which the student was encouraged to answer in writing a story about the picture. In Forms E, F, G, and J the same questions were contained in the instructions on the cover of the test booklets, but the remaining pages were left completely blank except for numbering. The format used in Form W may have had the effect of raising the average test score by eliciting more achievement imagery.

6 The 92 sets of data came from 36 pictures on high school subjects, 36 on three officer candidate groups, 12 on the fourth officer candidate group, and 8 on the college freshman group.

References

1. MacArthur, C. The effects of need achievement on the content of TAT stories: A re-examination. J. abnorm. soc. Psychol., 1953, 48, 532-536.

2. McClelland, D. C., Clark, R. A., Roby, T. B., and Atkinson, J. W. Projective expression of needs: IV. The effect of the need for achievement on thematic apperception. J. exp. Psychol., 1949, 39, 242-255.

3. McClelland, D. C., Clark, R. A., and Morgan, H. H. Revisions in the system of scoring imaginative stories for need achievement (nAch), scoring system D. Unpublished manuscript, April 1952.

4. McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.

5. Morgan, H. H. Measuring achievement motivation with "Picture Interpretations." J. consult. Psychol., 1953, 17, 289-292.

6. Ricciuti, H. N. and Clark, R. A. A comparison of need-achievement stories written by experimentally "relaxed" and "achievement-oriented" subjects: Effects obtained with new pictures and revised scoring categories. Princeton, N. J.: Educational Testing Service. (Technical Report No. 4, Office of Naval Research Contract Nonr-694(00), 1954.)

7. Ricciuti, H. N. and Sadacca, R. The prediction of academic grades with a projective test of achievement motivation: II. Cross-validation at the high school level. Princeton, N. J.: Educational Testing Service. (Technical Report No. 2, Office of Naval Research Contract Nonr-694(00), 1954.)

8. Sadacca, R., Ricciuti, H. N., and Swanson, E. Content analysis of need-achievement protocols: A study of scorer agreement. Princeton, N. J.: Educational Testing Service. (Technical Report No. 3, Office of Naval Research Contract Nonr-694(00), 1954.)